12-2012

Some theory for propensity-score-adjustment estimators in survey sampling

Jae Kwang Kim
Iowa State University, jkim@iastate.edu

Minsun Kim Riddles
Iowa State University

Survey Methodology, December 2012
Vol. 38, No. 2, pp. 157-165
Statistics Canada, Catalogue No. 12-001-X

Some theory for propensity-score-adjustment estimators in survey sampling

Jae Kwang Kim and Minsun Kim Riddles [1]

Abstract

The propensity-scoring-adjustment approach is commonly used to handle selection bias in survey sampling applications, including unit nonresponse and undercoverage. The propensity score is computed using auxiliary variables observed throughout the sample. We discuss some asymptotic properties of propensity-score-adjusted estimators and derive optimal estimators based on a regression model for the finite population. An optimal propensity-score-adjusted estimator can be implemented using an augmented propensity model. Variance estimation is discussed and the results from two simulation studies are presented.

Key Words: Calibration; Missing data; Nonresponse; Weighting.

1. Introduction

Consider a finite population of size N, where N is known. For each unit i, y_i is the study variable and $\mathbf{x}_i$ is the q-dimensional vector of auxiliary variables. The parameter of interest is the finite population mean of the study variable, $\theta = N^{-1}\sum_{i=1}^{N} y_i$. The finite population $\mathcal{F}_N = \{(\mathbf{x}_1', y_1), (\mathbf{x}_2', y_2), \ldots, (\mathbf{x}_N', y_N)\}$ is assumed to be a random sample of size N from a superpopulation distribution $F(\mathbf{x}, y)$. Suppose a sample of size n is drawn from the finite population according to a probability sampling design. Let $w_i = \pi_i^{-1}$ be the design weight, where π_i is the first-order inclusion probability of unit i obtained from the probability sampling design. Under complete response, the finite population mean can be estimated by the Horvitz-Thompson (HT) estimator, $\hat{\theta}_{\text{HT}} = N^{-1}\sum_{i \in A} w_i y_i$, where A is the set of indices appearing in the sample.

In the presence of missing data, the HT estimator $\hat{\theta}_{\text{HT}}$ cannot be computed. Let r be the response indicator variable that takes the value one if y is observed and takes the value zero otherwise. Conceptually, as discussed by Fay (1992), Shao and Steel (1999), and Kim and Rao (2009), the response indicator can be extended to the entire population as $\mathcal{R}_N = \{r_1, r_2, \ldots, r_N\}$, where r_i is a realization of the random variable r. In this case, the complete-case (CC) estimator $\hat{\theta}_{\text{CC}} = \sum_{i \in A} w_i r_i y_i / \sum_{i \in A} w_i r_i$ converges in probability to $E(Y \mid r = 1)$. Unless the response mechanism is missing completely at random in the sense that $E(Y \mid r = 1) = E(Y)$, the CC estimator is biased. To correct for the bias of the CC estimator, if the response probability

$$p(\mathbf{x}, y) = \Pr(r = 1 \mid \mathbf{x}, y) \tag{1}$$

is known, then the weighted CC estimator $\hat{\theta}_{\text{WCC}} = N^{-1}\sum_{i \in A} w_i r_i y_i / p(\mathbf{x}_i, y_i)$ can be used to estimate θ. Note that $\hat{\theta}_{\text{WCC}}$ is unbiased because $E\{\sum_{i \in A} w_i r_i y_i / p(\mathbf{x}_i, y_i) \mid \mathcal{F}_N\} = E\{\sum_{i=1}^{N} r_i y_i / p(\mathbf{x}_i, y_i) \mid \mathcal{F}_N\} = \sum_{i=1}^{N} y_i$.

If the response probability (1) is unknown, one can postulate a parametric model for the response probability $p(\mathbf{x}, y; \phi)$ indexed by $\phi \in \Omega$ such that $p(\mathbf{x}, y) = p(\mathbf{x}, y; \phi_0)$ for some $\phi_0 \in \Omega$. We assume that there exists a $\sqrt{n}$-consistent estimator $\hat{\phi}$ of ϕ_0 such that

$$\sqrt{n}(\hat{\phi} - \phi_0) = O_p(1), \tag{2}$$

where $g_n = O_p(1)$ indicates g_n is bounded in probability. Using $\hat{\phi}$, we can obtain the estimated response probability by $\hat{p}_i = p(\mathbf{x}_i, y_i; \hat{\phi})$, which is often called the propensity score (Rosenbaum and Rubin 1983). The propensity-score-adjusted (PSA) estimator can be constructed as

$$\hat{\theta}_{\text{PSA}} = \frac{1}{N}\sum_{i \in A} w_i \frac{r_i}{\hat{p}_i} y_i. \tag{3}$$

The PSA estimator (3) is widely used. Many surveys use the PSA estimator to reduce nonresponse bias (Fuller, Loughin and Baker 1994; Rizzo, Kalton and Brick 1996). Rosenbaum and Rubin (1983) and Rosenbaum (1987) proposed using the PSA approach to estimate the treatment effects in observational studies. Little (1988) reviewed the PSA methods for handling unit nonresponse in survey sampling. Duncan and Stasny (2001) used the PSA approach to control coverage bias in telephone surveys. Folsom (1991) and Iannacchione, Milne and Folsom (1991) used a logistic regression model for the response probability estimation. Lee (2006) applied the PSA method to a volunteer panel web survey. Durrant and Skinner (2006) used the PSA approach to address measurement error.

Despite the popularity of PSA estimators, asymptotic properties of PSA estimators have not received much attention in survey sampling literature. Kim and Kim (2007) used a Taylor expansion to obtain the asymptotic mean and variance of PSA estimators and discussed variance estimation. Da Silva and Opsomer (2006) and Da Silva and

1. Jae Kwang Kim and Minsun Kim Riddles, Department of Statistics, Iowa State University, Ames, IA, U.S.A. 50011. E-mail: jkim@iastate.edu.

Opsomer (2009) considered nonparametric methods to obtain PSA estimators.

In this paper, we discuss optimal PSA estimators in the class of PSA estimators of the form (3) that use a $\sqrt{n}$-consistent estimator $\hat{\phi}$. Such estimators are asymptotically unbiased for θ. Finding minimum variance PSA estimators among this particular class of PSA estimators is a topic of major interest in this paper.

Section 2 presents the main results. An optimal PSA estimator using an augmented propensity score model is proposed in Section 3. In Section 4, variance estimation of the proposed estimator is discussed. Results from two simulation studies can be found in Section 5 and concluding remarks are made in Section 6.

2. Main results

In this section, we discuss some asymptotic properties of PSA estimators. We assume that the response mechanism does not depend on y. Thus, we assume that

$$\Pr(r = 1 \mid \mathbf{x}, y) = \Pr(r = 1 \mid \mathbf{x}) = p(\mathbf{x}; \phi_0) \tag{4}$$

for some unknown vector ϕ_0. The first equality implies that the data are missing-at-random (MAR), as we always observe $\mathbf{x}$ in the sample. Note that the MAR condition is assumed in the population model. In the second equality, we further assume that the response mechanism is known up to an unknown parameter ϕ_0. The response mechanism is slightly different from that of Kim and Kim (2007), where the response mechanism is assumed to be under the classical two-phase sampling setup and depends on the realized sample:

$$\Pr(r = 1 \mid \mathbf{x}, y, I = 1) = \Pr(r = 1 \mid \mathbf{x}, I = 1) = p(\mathbf{x}; \phi_A^0). \tag{5}$$

Here, I is the sampling indicator function defined throughout the population. That is, $I_i = 1$ if $i \in A$ and $I_i = 0$ otherwise. Unless the sampling design is non-informative in the sense that the sample selection probabilities are correlated with the response indicator even after conditioning on auxiliary variables (Pfeffermann, Krieger and Rinott 1998), the two response mechanisms, (4) and (5), are different. In survey sampling, assumption (4) is more appropriate because an individual's decision on whether or not to respond to a survey is at his or her own discretion. Here, the response indicator variable r_i is defined throughout the population, as discussed in Section 1.

We consider a class of $\sqrt{n}$-consistent estimators of ϕ_0 in (4). In particular, we consider a class of estimators which can be written as a solution to

$$\hat{\mathbf{U}}_h(\phi) \equiv \sum_{i \in A} w_i \{ r_i - p_i(\phi) \} \mathbf{h}_i(\phi) = \mathbf{0}, \tag{6}$$

where $p_i(\phi) = p(\mathbf{x}_i; \phi)$ for some function $\mathbf{h}_i(\phi) = \mathbf{h}(\mathbf{x}_i; \phi)$, a smooth function of $\mathbf{x}_i$ and parameter ϕ. Thus, the solution to (6) can be written as $\hat{\phi}_h$, which depends on the choice of $\mathbf{h}_i(\phi)$. Any solution $\hat{\phi}_h$ to (6) is consistent for ϕ_0 in (4) because $E\{\hat{\mathbf{U}}_h(\phi_0) | \mathcal{F}_N\} = E[\sum_{i=1}^N \{r_i - p_i(\phi_0)\} \mathbf{h}_i(\phi_0) \mid \mathcal{F}_N]$ is zero under the response mechanism in (4). If we drop the sampling weights w_i in (6), the estimated parameter $\hat{\phi}_h$ is consistent for ϕ_A^0 in (5) and the resulting PSA estimator is consistent only when the sampling design is non-informative. The PSA estimators obtained from (6) using the sampling weights are consistent regardless of whether the sampling design is non-informative or not. According to Chamberlain (1987), any $\sqrt{n}$-consistent estimator of ϕ_0 in (4) can be written as a solution to (6). Thus, the choice of $\mathbf{h}_i(\phi)$ in (6) determines the efficiency of the resulting PSA estimator.

Let $\hat{\theta}_{\text{PSA},h}$ be the PSA estimator in (3) using $\hat{p}_i = p_i(\hat{\phi}_h)$ with $\hat{\phi}_h$ being the solution to (6). To discuss the asymptotic properties of $\hat{\theta}_{\text{PSA},h}$, assume a sequence of finite populations and samples, as in Isaki and Fuller (1982), such that $\sum_{i \in A} w_i \mathbf{u}_i - \sum_{i=1}^N \mathbf{u}_i = O_p(n^{-1/2} N)$ for any population characteristics $\mathbf{u}_i$ with bounded fourth moments. We also assume that the sampling weights are uniformly bounded. That is, $K_1 < N^{-1} n w_i < K_2$ for all i uniformly in n, where K_1 and K_2 are fixed constants. In addition, we assume the following regularity conditions:

[C1] The response mechanism satisfies (4), where $p(\mathbf{x}; \phi)$ is continuous in ϕ with continuous first and second derivatives in an open set containing ϕ_0. The responses are independent in the sense that $\text{Cov}(r_i, r_j \mid \mathbf{x}) = 0$ for $i \neq j$. Also, $p(\mathbf{x}_i; \phi) > c$ for all i for some fixed constant $c > 0$.

[C2] The solution to (6) exists and is unique almost everywhere. The function $\mathbf{h}_i(\phi) = \mathbf{h}(\mathbf{x}_i; \phi)$ in (6) has a bounded fourth moment. Furthermore, the partial derivative $\partial \{\hat{\mathbf{U}}_h(\phi)\} / \partial \phi$ is nonsingular for all n.

[C3] The estimating function $\hat{\mathbf{U}}_h(\phi)$ in (6) converges in probability to $\mathbf{U}_h(\phi) = \sum_{i=1}^N \{r_i - p_i(\phi)\} \mathbf{h}_i(\phi)$ uniformly in ϕ. Furthermore, the partial derivative $\partial \{\hat{\mathbf{U}}_h(\phi)\} / \partial \phi$ converges in probability to $\partial \{\mathbf{U}_h(\phi)\} / \partial \phi$ uniformly in ϕ. The solution ϕ_N to $\mathbf{U}_h(\phi) = \mathbf{0}$ satisfies $N^{1/2}(\phi_N - \phi_0) = O_p(1)$ under the response mechanism.

Condition [C1] states the regularity conditions for the response mechanism. Condition [C2] is the regularity condition for the solution $\hat{\phi}_h$ to (6). In Condition [C3], some regularity conditions are imposed on the estimating function $\hat{\mathbf{U}}_h(\phi)$ itself. By [C2] and [C3], we can establish the $\sqrt{n}$-consistency (2) of $\hat{\phi}_h$.

Now, the following theorem deals with some asymptotic properties of the PSA estimator $\hat{\theta}_{\text{PSA},\,h}$.

Theorem 1 If conditions [C1] - [C3] hold, then under the joint distribution of the sampling mechanism and the response mechanism, the PSA estimator $\hat{\theta}_{\text{PSA},\,h}$ *satisfies*

$$\sqrt{n}\,(\hat{\theta}_{\text{PSA},\,h} - \tilde{\theta}_{\text{PSA},\,h}) = o_p(1), \tag{7}$$

where

$$\tilde{\theta}_{\text{PSA},\,h} = \frac{1}{N}\sum_{i\in A} w_i \left\{ p_i\,\mathbf{h}_i'\,\gamma_h^* + \frac{r_i}{p_i}(y_i - p_i\,\mathbf{h}_i'\,\gamma_h^*)\right\}, \tag{8}$$

$\gamma_h^* = (\sum_{i=1}^N r_i\,\mathbf{z}_i\,p_i\,\mathbf{h}_i')^{-1}(\sum_{i=1}^N r_i\,\mathbf{z}_i\,y_i)$, $p_i = p(\boldsymbol{x}_i;\phi_0)$, $\mathbf{z}_i = \partial\{p^{-1}(\boldsymbol{x}_i;\phi_0)\}/\partial\phi$, *and* $\mathbf{h}_i = \mathbf{h}(\boldsymbol{x}_i;\phi_0)$. *Moreover, if the finite population is a random sample from a superpopulation model, then*

$$V(\tilde{\theta}_{\text{PSA},h}) \geq V_l \equiv V(\hat{\theta}_{\text{HT}}) + \frac{1}{N^2}E\left\{\sum_{i\in A} w_i^2\left(\frac{1}{p_i}-1\right)V(Y\mid \boldsymbol{x}_i)\right\}. \tag{9}$$

The equality in (9) holds when $\hat{\phi}_h$ *satisfies*

$$\sum_{i\in A} w_i\left\{\frac{r_i}{p(\boldsymbol{x}_i;\hat{\phi}_h)} - 1\right\}E(Y\mid \boldsymbol{x}_i) = 0, \tag{10}$$

where $E(Y\mid \boldsymbol{x}_i)$ *is the conditional expectation under the superpopulation model.*

Proof. Given $p_i(\phi) = p(\boldsymbol{x}_i;\phi)$ *and* $\mathbf{h}_i(\phi) = \mathbf{h}(\boldsymbol{x}_i;\phi)$, *define*

$$\hat{\theta}(\phi,\gamma) = N^{-1}\sum_{i\in A} w_i\left[p_i(\phi)\mathbf{h}_i'(\phi)\gamma + \frac{r_i}{p_i(\phi)}\{y_i - p_i(\phi)\mathbf{h}_i'(\phi)\,\gamma\}\right].$$

Since $\hat{\phi}_h$ *satisfies (6), we have* $\hat{\theta}_{\text{PSA}} = \hat{\theta}(\hat{\phi}_h,\gamma)$ *for any choice of* γ. *We now want to find a particular choice of* γ, *say* γ^*, *such that*

$$\hat{\theta}(\hat{\phi}_h,\gamma^*) = \hat{\theta}(\phi_0,\gamma^*) + o_p(n^{-1/2}). \tag{11}$$

As $\hat{\phi}_h$ *converges in probability to* ϕ_0, *the asymptotic equivalence (11) holds if*

$$E\left\{\frac{\partial}{\partial\phi}\hat{\theta}(\phi,\gamma^*)\mid \phi = \phi_0\right\} = \mathbf{0}, \tag{12}$$

using the theory of Randles (1982). Condition (12) holds if $\gamma^* = \gamma_h^*$, *where* γ_h^* *is defined in (8). Thus, (11) reduces to*

$$\hat{\theta}_{\text{PSA},\,h} = \frac{1}{N}\sum_{i\in A} w_i\left\{p_i\,\mathbf{h}_i'\,\gamma_h^* + \frac{r_i}{p_i}(y_i - p_i\,\mathbf{h}_i'\,\gamma_h^*)\right\} + o_p(n^{-1/2}), \tag{13}$$

which proves (7). The variance of $\tilde{\theta}_{\text{PSA},\,h}$ *can be derived as*

$$\begin{aligned}
V(\tilde{\theta}_{\text{PSA},\,h}) &= V(\hat{\theta}_{\text{HT}}) + \frac{1}{N^2}E\left\{\sum_{i\in A} w_i^2\left(\frac{1}{p_i}-1\right)(y_i - p_i\mathbf{h}_i'\gamma_h^*)^2\right\}\\
&= V(\hat{\theta}_{\text{HT}}) + \frac{1}{N^2}E\left[\sum_{i\in A} w_i^2\left(\frac{1}{p_i}-1\right)\{y_i - E(Y\mid \boldsymbol{x}_i) + E(Y\mid \boldsymbol{x}_i) - p_i\mathbf{h}_i'\gamma_h^*\}^2\right]\\
&= V(\hat{\theta}_{\text{HT}}) + \frac{1}{N^2}E\left\{\sum_{i\in A} w_i^2\left(\frac{1}{p_i}-1\right)V(Y\mid \boldsymbol{x}_i)\right\}\\
&\quad + \frac{1}{N^2}E\left[\sum_{i\in A} w_i^2\left(\frac{1}{p_i}-1\right)\{E(Y\mid \boldsymbol{x}_i) - p_i\mathbf{h}_i'\gamma_h^*\}^2\right],
\end{aligned} \tag{14}$$

where the last equality follows because y_i *is conditionally independent of* $E(Y\mid \boldsymbol{x}_i) - p_i\mathbf{h}_i'\gamma_h^*$, *conditioning on* $\mathbf{x}_i$. *Since the last term in (14) is non-negative, the inequality in (9) is established. Furthermore, if* $E(Y\mid \boldsymbol{x}_i) = p_i\,\mathbf{h}_i'\boldsymbol{\alpha}$ *for some* $\boldsymbol{\alpha}$, *then (10) holds and* $E(\gamma_h^*\mid \boldsymbol{x}_i) = \boldsymbol{\alpha}$, *by the definition of* γ_h^*. *Thus,* $E(Y\mid \boldsymbol{x}_i) - p_i\mathbf{h}_i'\gamma_h^* = -p_i\mathbf{h}_i'\{\gamma_h^* - E(\gamma_h^*\mid \boldsymbol{x}_i)\} = o_p(1)$, *implying that the last term in (14) is negligible.*

In (9), V_l is the lower bound of the asymptotic variance of PSA estimators of the form (3) satisfying (6). Any PSA estimator that has the asymptotic variance V_l in (9) is optimal in the sense that it achieves the lower bound of the asymptotic variance among the class of PSA estimators with $\hat{\phi}$ satisfying (2). The asymptotic variance of optimal PSA estimators of θ is equal to V_l in (9). The PSA estimator using the maximum likelihood estimator of ϕ_0 does not necessarily achieve the lower bound of the asymptotic variance.

Condition (10) provides a way of constructing an optimal PSA estimator. First, we need an assumption for $E(Y\mid \mathbf{x})$, which is often called the outcome regression model. If the outcome regression model is a linear regression model of the form $E(Y\mid \mathbf{x}) = \beta_0 + \boldsymbol{\beta}_1'\mathbf{x}$, an optimal PSA estimator of θ can be obtained by solving

$$\sum_{i\in A} w_i\,\frac{r_i}{p_i(\phi)}(1,\mathbf{x}_i) = \sum_{i\in A} w_i\,(1,\mathbf{x}_i). \tag{15}$$

Condition (15) is appealing because it says that the PSA estimator applied to $y = a + \mathbf{b}'\mathbf{x}$ leads to the original HT estimator. Condition (15) is called the calibration condition in survey sampling. The calibration condition applied to $\mathbf{x}$ makes full use of the information contained in it if the study variable is well approximated by a linear function of $\mathbf{x}$. Condition (15) was also used in Nevo (2003) and Kott (2006) under the linear regression model.

If we explicitly use a regression model for $E(Y \mid \mathbf{x})$, it is possible to construct an estimator that has asymptotic variance (9) and is not necessarily a PSA estimator. For example, if we assume that

$$E(Y \mid \mathbf{x}) = m(\mathbf{x}; \beta_0) \tag{16}$$

for some function $m(\mathbf{x}; \cdot)$ known up to β_0, we can use the model (16) directly to construct an optimal estimator of the form

$$\hat{\theta}_{\text{opt}} = \frac{1}{N} \sum_{i \in A} w_i \left[m(\mathbf{x}_i; \hat{\beta}) + \frac{r_i}{p_i(\hat{\phi})} \{ y_i - m(\mathbf{x}_i; \hat{\beta}) \} \right], \tag{17}$$

where $\hat{\beta}$ is a $\sqrt{n}$-consistent estimator of β_0 in the superpopulation model (16) and $\hat{\phi}$ is a $\sqrt{n}$-consistent estimator of ϕ_0 computed by (6). The following theorem shows that the optimal estimator (17) achieves the lower bound in (9).

Theorem 2 Let the conditions of Theorem 1 hold. Assume that $\hat{\beta}$ satisfies $\hat{\beta} = \beta_0 + O_p(n^{-1/2})$. Assume that, in the superpopulation model (16), $m(\boldsymbol{x}; \beta)$ has continuous first-order partial derivatives in an open set containing β_0. Under the joint distribution of the sampling mechanism, the response mechanism, and the superpopulation model (16), the estimator $\hat{\theta}_{\text{opt}}$ in (17) satisfies

$$\sqrt{n}\,(\hat{\theta}_{\text{opt}} - \tilde{\theta}^*_{\text{opt}}) = o_p(1),$$

where

$$\tilde{\theta}^*_{\text{opt}} = N^{-1} \sum_{i \in A} w_i \left[m(\boldsymbol{x}_i; \beta_0) + \frac{r_i}{p_i} \{ y_i - m(\boldsymbol{x}_i; \beta_0) \} \right],$$

$p_i = p_i(\phi_0)$, and $V(\tilde{\theta}^*_{\text{opt}})$ is equal to V_l in (9).

Proof. Define $\hat{\theta}_{\text{opt}}(\beta, \phi) = N^{-1} \sum_{i \in A} w_i [m(\boldsymbol{x}_i; \beta) + r_i\, p_i^{-1}(\phi) \{y_i - m(\boldsymbol{x}_i; \beta)\}]$. Note that $\hat{\theta}_{\text{opt}}$ in (17) can be written as $\hat{\theta}_{\text{opt}} = \hat{\theta}_{\text{opt}}(\hat{\beta}, \hat{\phi})$. Since

$$\frac{\partial}{\partial \beta} \hat{\theta}_{\text{opt}}(\beta, \phi) = \frac{1}{N} \sum_{i \in A} w_i \left\{ \breve{m}(\boldsymbol{x}_i; \beta) - \frac{r_i}{p_i(\phi)} \breve{m}(\boldsymbol{x}_i; \beta) \right\},$$

where $\breve{m}(\boldsymbol{x}_i; \beta) = \partial m(\boldsymbol{x}_i; \beta) / \partial \beta$, and

$$\frac{\partial}{\partial \phi} \hat{\theta}_{\text{opt}}(\beta, \phi) = \frac{1}{N} \sum_{i \in A} w_i\, r_i\, \mathbf{z}_i(\phi) \{ y_i - m(\boldsymbol{x}_i; \beta) \},$$

where $\mathbf{z}_i(\phi) = \partial \{ p_i^{-1}(\phi) \} / \partial \phi$, we have $E[\partial \{ \hat{\theta}_{\text{opt}}(\beta, \phi) \} / \partial(\beta, \phi) \mid \beta = \beta_0, \phi = \phi_0] = \mathbf{0}$ and the condition of Randles (1982) is satisfied. Thus,

$$\hat{\theta}_{\text{opt}}(\hat{\beta}, \hat{\phi}) = \hat{\theta}_{\text{opt}}(\beta_0, \phi_0) + o_p(n^{-1/2}) = \tilde{\theta}^*_{\text{opt}} + o_p(n^{-1/2})$$

and the variance of $\tilde{\theta}^*_{\text{opt}}$ is equal to V_l, the lower bound of the asymptotic variance.

The (asymptotic) optimality of the estimator in (17) is justified under the joint distribution of the response model (4) and the superpopulation model (16). When both models are correct, $\hat{\theta}_{\text{opt}}$ is optimal and the choice of $(\hat{\beta}, \hat{\phi})$ does not affect the efficiency of the $\hat{\theta}_{\text{opt}}$ as long as $(\hat{\beta}, \hat{\phi})$ is $\sqrt{n}$-consistent. Robins, Rotnitzky and Zhao (1994) also advocated using $\hat{\theta}_{\text{opt}}$ in (17) under simple random sampling.

Remark 1 When the response model is correct and the superpopulation model (16) is not necessarily correct, the choice of $\hat{\beta}$ does affect the efficiency of the optimal estimator. Cao, Tsiatis and Davidian (2009) considered optimal estimation when only the response model is correct. Using Taylor linearization, the optimal estimator in (17) with $\hat{\phi}$ satisfying (6) is asymptotically equivalent to

$$\tilde{\theta}(\beta) = \sum_{i \in A} w_i \left[m(\boldsymbol{x}_i; \beta) + \frac{r_i}{p_i} \{ y_i - m(\boldsymbol{x}_i; \beta) \} - \left(\frac{r_i}{p_i} - 1 \right) \mathbf{c}'_\beta\, p_i\, \mathbf{h}_i \right],$$

where $\mathbf{c}_\beta$ is the probability limit of $\hat{\mathbf{c}}_\beta = \{ \sum_{i \in A} w_i\, r_i\, \mathbf{z}_i(\hat{\phi})\, \hat{p}_i \mathbf{h}'_i(\hat{\phi}) \}^{-1} \sum_{i \in A} w_i\, r_i\, \mathbf{z}_i(\hat{\phi}) \{ y_i - m(\boldsymbol{x}_i; \beta) \}$ and $\mathbf{z}_i(\phi) = \partial \{ p_i^{-1}(\phi) \} / \partial \phi$. The asymptotic variance is then equal to

$$V\{\tilde{\theta}(\beta)\} = V(\hat{\theta}_{\text{HT}}) + E\left[\sum_{i \in A} w_i^2 \frac{1 - p_i}{p_i} \{ y_i - m(\boldsymbol{x}_i; \beta) - \mathbf{c}'_\beta\, p_i\, \mathbf{h}_i \}^2 \right].$$

Thus, an optimal estimator of β can be computed by finding $\hat{\beta}$ that minimizes

$$Q(\beta) = \sum_{i \in A} w_i^2\, r_i \frac{1 - \hat{p}_i}{\hat{p}_i^2} \{ y_i - m(\boldsymbol{x}_i; \beta) - \hat{\mathbf{c}}'_\beta\, \hat{p}_i\, \mathbf{h}_i(\hat{\phi}) \}^2.$$

The resulting estimator is design-optimal in the sense that it minimizes the asymptotic variance under the response model.

3. Augmented propensity score model

In this section, we consider optimal PSA estimation. Note that the optimal estimator $\hat{\theta}_{\text{opt}}$ in (17) is not necessarily written as a PSA estimator form in (3). It is in the PSA estimator form if it satisfies $\sum_{i \in A} w_i\, r_i\, \hat{p}_i^{-1} m(\mathbf{x}_i; \hat{\beta}) = \sum_{i \in A} w_i\, m(\mathbf{x}_i; \hat{\beta})$. Thus, we can construct an optimal PSA estimator by including $m(\mathbf{x}_i; \hat{\beta})$ in the model for the propensity score. Specifically, given $\hat{m}_i = m(\mathbf{x}_i; \hat{\beta})$, $\hat{p}_i = p_i(\hat{\phi})$ and $\hat{\mathbf{h}}_i = \mathbf{h}_i(\hat{\phi})$, where $\hat{\phi}$ is obtained from (6), we augment the response model by

$$p_i^*(\hat{\phi}, \lambda) \equiv \frac{\hat{p}_i}{\hat{p}_i + (1 - \hat{p}_i) \exp(\lambda_0 + \lambda_1 \hat{m}_i)}, \tag{18}$$

where $\lambda = (\lambda_0, \lambda_1)'$ is the Lagrange multiplier which is used to incorporate the additional constraint. If $(\lambda_0, \lambda_1)' = \mathbf{0}$, then $p_i^*(\hat{\phi}, \lambda) = \hat{p}_i$. The augmented response probability $p_i^*(\hat{\phi}, \lambda)$ always takes values between 0 and 1. The augmented response probability model (18) can be derived by minimizing the Kullback-Leibler distance $\sum_{i \in A} w_i\, r_i\, q_i^* \log(q_i^* / q_i)$, where $q_i^* = (1 - p_i^*)/p_i^*$ and $q_i = (1 - \hat{p}_i)/\hat{p}_i$, subject to the constraint $\sum_{i \in A} w_i (r_i / p_i^*)(1, \hat{m}_i) = \sum_{i \in A} w_i (1, \hat{m}_i)$.

Using (18), the optimal PSA estimator is computed by

$$\hat{\theta}^*_{\text{PSA}} = \frac{1}{N} \sum_{i \in A} w_i \frac{r_i}{p_i^*(\hat{\phi}, \hat{\lambda})} y_i, \tag{19}$$

where $\hat{\lambda}$ satisfies

$$\sum_{i \in A} w_i \frac{r_i}{p_i^*(\hat{\phi}, \hat{\lambda})} (1, \hat{m}_i) = \sum_{i \in A} w_i (1, \hat{m}_i). \tag{20}$$

Under the response model (4), it can be shown that

$$\hat{\theta}^*_{\text{PSA}} = \frac{1}{N} \sum_{i \in A} w_i \left\{ \hat{b}_0 + \hat{b}_1 \hat{m}_i + \frac{r_i}{\hat{p}_i} (y_i - \hat{b}_0 - \hat{b}_1 \hat{m}_i) \right\} + o_p(n^{-1/2}),$$

where

$$\begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \end{pmatrix} = \left\{ \sum_{i \in A} w_i\, r_i \left(\frac{1}{\hat{p}_i} - 1 \right) \begin{pmatrix} 1 \\ \hat{m}_i \end{pmatrix} \begin{pmatrix} 1 \\ \hat{m}_i \end{pmatrix}' \right\}^{-1} \sum_{i \in A} w_i\, r_i \left(\frac{1}{\hat{p}_i} - 1 \right) \begin{pmatrix} 1 \\ \hat{m}_i \end{pmatrix} y_i. \tag{21}$$

Furthermore, by the argument for Theorem 1, we can establish that

$$\begin{aligned} \hat{\theta}^*_{\text{PSA}} = \frac{1}{N} \sum_{i \in A} w_i \Big\{ & b_0 + b_1 \hat{m}_i + \gamma'_{h2}\, p_i\, \mathbf{h}_i \\ & + \frac{r_i}{p_i} \left(y_i - b_0 - b_1 \hat{m}_i - \gamma'_{h2}\, p_i\, \mathbf{h}_i \right) \Big\} \\ & + o_p(n^{-1/2}), \end{aligned}$$

where (b_0, b_1, γ'_{h2}) is the probability limit of $(\hat{b}_0, \hat{b}_1, \hat{\gamma}'_{h2})$ with

$$\hat{\gamma}_{h2} = \left\{ \sum_{i \in A} w_i\, r_i\, \mathbf{z}_i(\hat{\phi})\, \hat{p}_i\, \mathbf{h}'_i(\hat{\phi}) \right\}^{-1} \sum_{i \in A} w_i\, r_i\, \mathbf{z}_i(\hat{\phi}) (y_i - \hat{b}_0 - \hat{b}_1 \hat{m}_i) \tag{22}$$

and the effect of estimating ϕ_0 in $\hat{p}_i = p(\mathbf{x}_i; \hat{\phi})$ can be safely ignored.

Note that, under the response model (4), $(\hat{\phi}, \hat{\lambda})$ in (19) converges in probability to $(\phi_0, \mathbf{0})$, where ϕ_0 is the true parameter in (4). Thus, the propensity score from the augmented model converges to the true response probability. Because $\hat{\lambda}$ converges to zero in probability, the choice of $\hat{\beta}$ in $\hat{m}_i = m(x_i; \hat{\beta})$ does not play a role for the asymptotic unbiasedness of the PSA estimator. The asymptotic variances are changed for different choices of $\hat{\beta}$.

Under the superpopulation model (16), $\hat{b}_0 + \hat{b}_1 \hat{m}_i \to E(Y \mid \mathbf{x}_i)$ in probability. Thus, the optimal PSA estimator in (19) is asymptotically equivalent to the optimal estimator in (17). Incorporating $\hat{m}_i$ into the calibration equation to achieve optimality is close in spirit to the model-calibration method proposed by Wu and Sitter (2001).

4. Variance estimation

We now discuss variance estimation of PSA estimators under the assumed response model. Singh and Folsom (2000) and Kott (2006) discussed variance estimation for certain types of PSA estimators. Kim and Kim (2007) discussed variance estimation when the PSA estimator is computed with the maximum likelihood method.

We consider variance estimation for the PSA estimator of the form (3) where $\hat{p}_i = p_i(\hat{\phi})$ is constructed to satisfy (6) for some $\mathbf{h}_i(\phi) = \mathbf{h}(\mathbf{x}_i; \phi, \hat{\beta})$, where $\hat{\beta}$ is obtained using the postulated superpopulation model. Let β^* be the probability limit of $\hat{\beta}$ under the response model. Note that β^* is not necessarily equal to β_0 in (16) since we are not assuming that the postulated superpopulation model is correctly specified in this section.

Using the argument for the Taylor linearization (13) used in the proof of Theorem 1, the PSA estimator satisfies

$$\hat{\theta}_{\text{PSA}} = \frac{1}{N} \sum_{i \in A} w_i\, \eta_i(\phi_0, \beta^*) + o_p(n^{-1/2}), \tag{23}$$

where

$$\begin{aligned} \eta_i(\phi, \beta) = {} & p_i(\phi) \mathbf{h}'_i(\phi, \beta)\, \gamma_h^* \\ & + \frac{r_i}{p_i(\phi)} \{ y_i - p_i(\phi) \mathbf{h}'_i(\phi, \beta)\, \gamma_h^* \}, \end{aligned} \tag{24}$$

$\mathbf{h}_i(\phi, \beta) = \mathbf{h}(\mathbf{x}_i; \phi, \beta)$ and γ_h^* is defined as in (8) with $\mathbf{h}_i$ replaced by $\mathbf{h}_i(\phi_0, \beta^*)$. Since $p_i(\hat{\phi})$ satisfies (6) with $\mathbf{h}_i(\phi) = \mathbf{h}(\mathbf{x}_i; \phi, \hat{\beta})$, $\hat{\theta}_{\text{PSA}} = N^{-1} \sum_{i \in A} w_i\, \eta_i(\hat{\phi}, \hat{\beta})$ holds and the linearization in (23) can be expressed as $N^{-1} \sum_{i \in A} w_i\, \eta_i(\hat{\phi}, \hat{\beta}) = N^{-1} \sum_{i \in A} w_i\, \eta_i(\phi_0, \beta^*) + o_p(n^{-1/2})$. Thus, if $(\mathbf{x}_i, y_i, r_i)$ are independent and identically distributed (IID), then $\eta_i(\phi_0, \beta^*)$ are IID even though $\eta_i(\hat{\phi}, \hat{\beta})$ are not necessarily IID. Because $\eta_i(\phi_0, \beta^*)$ are IID, we can apply the standard complete sample method to estimate the variance of $\hat{\eta}_{\text{HT}} = N^{-1} \sum_{i \in A} w_i\, \eta_i(\phi_0, \beta^*)$, which is asymptotically equivalent to the variance of $\hat{\theta}_{\text{PSA}} = N^{-1} \sum_{i \in A} w_i\, \eta_i(\hat{\phi}, \hat{\beta})$. See Kim and Rao (2009).

To derive the variance estimator, we assume that the variance estimator $\hat{V} = N^{-2} \sum_{i \in A} \sum_{j \in A} \Omega_{ij}\, g_i\, g_j$ satisfies

$\hat{V}/V(\hat{g}_{\mathrm{HT}} \mid \mathcal{F}_N) = 1 + o_p(1)$ for some Ω_{ij} related to the joint inclusion probability, where $\hat{g}_{\mathrm{HT}} = N^{-1}\sum_{i\in A} w_i\, g_i$ for any g with a finite second moment and $V(g_{\mathrm{HT}} \mid \mathcal{F}_N) = N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Omega_{N\cdot ij}\, g_i\, g_j$, for some $\Omega_{N\cdot ij}$. We also assume that

$$\sum_{i=1}^{N} |\Omega_{N\cdot ij}| = O(n^{-1}N). \tag{25}$$

To obtain the total variance, the *reverse framework* of Fay (1992), Shao and Steel (1999), and Kim and Rao (2009) is considered. In this framework, the finite population is first divided into two groups, a population of respondents and a population of nonrespondents. Given the population, the sample A is selected according to a probability sampling design. Thus, selection of the population respondents from the whole finite population is treated as the first-phase sampling and the selection of the sample respondents from the population respondents is treated as the second-phase sampling in the reverse framework. The total variance of $\hat{\eta}_{\mathrm{HT}}$ can be written as

$$V(\hat{\eta}_{\mathrm{HT}} \mid \mathcal{F}_N) = V_1 + V_2 = E\{V(\hat{\eta}_{\mathrm{HT}} \mid \mathcal{F}_N, \mathcal{R}_N) \mid \mathcal{F}_N\} + V\{E(\hat{\eta}_{\mathrm{HT}} \mid \mathcal{F}_N, \mathcal{R}_N) \mid \mathcal{F}_N\}. \tag{26}$$

The conditional variance term $V(\hat{\eta}_{\mathrm{HT}} \mid \mathcal{F}_N, \mathcal{R}_N)$ in (26) can be estimated by

$$\hat{V}_1 = N^{-2}\sum_{i\in A}\sum_{j\in A}\Omega_{ij}\,\hat{\eta}_i\,\hat{\eta}_j, \tag{27}$$

where $\hat{\eta}_i = \eta_i(\hat{\phi}, \hat{\beta})$ is defined in (24) with γ_h^* replaced by a consistent estimator such as $\hat{\gamma}_h^* = \{\sum_{i\in A} w_i\, r_i\, \mathbf{z}_i(\hat{\phi})\,\hat{p}_i\, \hat{\mathbf{h}}_i'\}^{-1}\sum_{i\in A} w_i\, r_i\, \mathbf{z}_i(\hat{\phi})\, y_i$, and $\hat{\mathbf{h}}_i = \mathbf{h}(\mathbf{x}_i; \hat{\phi}, \hat{\beta})$. To show that $\hat{V}_1$ is also consistent for V_1 in (26), it suffices to show that $V\{n\cdot V(\hat{\eta}_{\mathrm{HT}} \mid \mathcal{F}_N, \mathcal{R}_N) \mid \mathcal{F}_N\} = o(1)$, which follows by (25) and the existence of the fourth moment. See Kim, Navarro and Fuller (2006). The second term V_2 in (26) is

$$\begin{aligned} V\{E(\hat{\eta}_{\mathrm{HT}} \mid \mathcal{F}_N, \mathcal{R}_N) \mid \mathcal{F}_N\} &= V\left(N^{-1}\sum_{i=1}^{N}\eta_i \mid \mathcal{F}_N\right) \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\frac{1-p_i}{p_i}\,(y_i - p_i\mathbf{h}_i^{*\prime}\gamma_h^*)^2, \end{aligned}$$

where $\mathbf{h}_i^* = \mathbf{h}(\mathbf{x}_i; \phi_0, \beta^*)$. A consistent estimator of V_2 can be derived as

$$\hat{V}_2 = \frac{1}{N^2}\sum_{i\in A} w_i\, r_i\,\frac{1-\hat{p}_i}{\hat{p}_i^2}\,(y_i - \hat{p}_i\hat{\mathbf{h}}_i'\,\hat{\gamma}_h^*)^2, \tag{28}$$

where $\hat{\gamma}_h^*$ is defined after (27). Therefore,

$$\hat{V}(\hat{\theta}_{\mathrm{PSA}}) = \hat{V}_1 + \hat{V}_2, \tag{29}$$

is consistent for the variance of the PSA estimator defined in (3) with $\hat{p}_i = p_i(\hat{\phi})$ satisfying (6), where $\hat{V}_1$ is in (27) and $\hat{V}_2$ is in (28).

Note that the first term of the total variance is $V_1 = O_p(n^{-1})$, but the second term is $V_2 = O_p(N^{-1})$. Thus, when the sampling fraction nN^{-1} is negligible, that is, $nN^{-1} = o(1)$, the second term V_2 can be ignored and $\hat{V}_1$ is a consistent estimator of the total variance. Otherwise, the second term V_2 should be taken into consideration, so that a consistent variance estimator can be constructed as in (29).

Remark 2 The variance estimation of the optimal PSA estimator with augmented propensity model (18) with $(\hat{\phi}, \hat{\lambda})$ *satisfying (20) can be derived by (29) using* $\hat{\eta}_i = \hat{b}_0 + \hat{b}_1\hat{m}_i + \hat{\gamma}_{h2}'\hat{p}_i\hat{\mathbf{h}}_i + r_i\hat{p}_i^{-1}(y_i - \hat{b}_0 - \hat{b}_1\hat{m}_i - \hat{\gamma}_{h2}'\hat{p}_i\hat{\mathbf{h}}_i)$ *where* $(\hat{b}_0, \hat{b}_1)$ *and* $\hat{\gamma}_{h2}$ *are defined in (21) and (22), respectively.*

5. Simulation study

5.1 Study one

Two simulation studies were performed to investigate the properties of the proposed method. In the first simulation, we generated a finite population of size $N = 10{,}000$ from the following multivariate normal distribution:

$$\begin{pmatrix} x_1 \\ x_2 \\ e \end{pmatrix} \sim N\left[\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right].$$

The variable of interest y was constructed as $y = 1 + x_1 + e$. We also generated response indicator variables r_i independently from a Bernoulli distribution with probability

$$p_i = \frac{\exp(2 + x_{2i})}{1 + \exp(2 + x_{2i})}.$$

From the finite population, we used simple random sampling to select two samples of size, $n = 100$ and $n = 400$, respectively. We used $B = 5{,}000$ Monte Carlo samples in the simulation. The average response rate was about 69.6%.

To compute the propensity score, a response model of the form

$$p(\mathbf{x}; \phi) = \frac{\exp(\phi_0 + \phi_1 x_2)}{1 + \exp(\phi_0 + \phi_1 x_2)} \tag{30}$$

was postulated and an outcome regression model of the form

$$m(\mathbf{x}; \beta) = \beta_0 + \beta_1 x_1 \tag{31}$$

was postulated to obtain the optimal PSA estimators. Thus, both models are correctly specified. From each sample, we computed four estimators of $\theta = N^{-1}\sum_{i=1}^{N} y_i$:

1. (PSA-MLE): PSA estimator in (3) with $\hat{p}_i = p_i(\hat{\phi})$ and $\hat{\phi}$ being the maximum likelihood estimator of ϕ.
2. (PSA-CAL): PSA estimator in (3) with $\hat{p}_i$ satisfying the calibration constraint (15) on $(1, x_{2i})$.
3. (AUG): Augmented PSA estimator in (19).
4. (OPT): Optimal estimator in (17).

In the augmented PSA estimators, $\hat{\phi}$ was computed by the maximum likelihood method. Under model (30), the maximum likelihood estimator of $\phi = (\phi_0, \phi_1)'$ was computed by solving (6) with $\mathbf{h}_i(\phi) = (1, x_{2i})'$. Parameter (β_0, β_1) for the outcome regression model was computed using ordinary least squares, regressing y on x_1. In addition to the point estimators, we also computed the variance estimators of the point estimators. The variance estimators of the PSA estimators were computed using the pseudo-values in (24) and the $\mathbf{h}_i(\phi)$ corresponding to each estimator. For the augmented PSA estimators, the pseudo-values were computed by the method in Remark 2.

Table 1 presents the Monte Carlo biases, variances, and mean square errors of the four point estimators and the Monte Carlo percent relative biases and t-statistics of the variance estimators of the estimators. The percent relative bias of a variance estimator $\hat{V}(\hat{\theta})$ is calculated as $100 \times \{V_{\mathrm{MC}}(\hat{\theta})\}^{-1} [E_{\mathrm{MC}}\{\hat{V}(\hat{\theta})\} - V_{\mathrm{MC}}(\hat{\theta})]$, where $E_{\mathrm{MC}}(\cdot)$ and $V_{\mathrm{MC}}(\cdot)$ denote the Monte Carlo expectation and the Monte Carlo variance, respectively. The t-statistic in Table 1 is the test statistic for testing the zero bias of the variance estimator. See Kim (2004).

Based on the simulation results in Table 1, we have the following conclusions.

1. All of the PSA estimators are asymptotically unbiased because the response model (30) is correctly specified. The PSA estimator using the calibration method is slightly more efficient than the PSA estimator using the maximum likelihood estimator, because the last term of (14) is smaller for the calibration method as the predictor for $E(Y \mid \mathbf{x}_i) = \beta_0 + \beta_1 x_{1i}$ is better approximated by a linear function of $(1, x_{2i})$ than by a linear function of $(\hat{p}_i, \hat{p}_i x_{2i})$.
2. The augmented PSA estimator is more efficient than the direct PSA estimator (3). The augmented PSA estimator is constructed by using the correctly specified regression model (31) and so it is asymptotically equivalent to the optimal PSA estimator in (17).
3. Variance estimators are all approximately unbiased. There are some modest biases in the variance estimators of the PSA estimators when the sample size is small $(n = 100)$.

5.2 Study two

In the second simulation study, we further investigated the PSA estimators with a non-linear outcome regression model under an unequal probability sampling design. We generated two stratified finite populations of (x, y) with four strata $(h = 1, 2, 3, 4)$, where x_{hi} were independently generated from a normal distribution $N(1, 1)$ and y_{hi} were dichotomous variables that take values of 1 or 0 from a Bernoulli distribution with probability p_{1yhi} or p_{2yhi}. Two different probabilities were used for two populations, respectively:

1. Population 1 (Pop1):

$$p_{1yhi} = 1/\{1 + \exp(0.5 - 2x)\}.$$

2. Population 2 (Pop2):

$$p_{2yhi} = 1 / [1 + \exp\{0.25(x - 1.5)^2 - 1.5\}].$$

In addition to x_{hi} and y_{hi}, the response indicator variables r_{hi} were generated from a Bernoulli distribution with probability $p_{hi} = 1 / \{1 + \exp(-1.5 + 0.7x_{hi})\}$. The sizes of the four strata were $N_1 = 1{,}000$, $N_2 = 2{,}000$, $N_3 = 3{,}000$, and $N_4 = 4{,}000$, respectively. In each of the two sets of finite population, a stratified sample of size $n = 400$ was independently generated without replacement, where a simple random sample of size $n_h = 100$ was selected from each stratum. We used $B = 5{,}000$ Monte Carlo samples in this simulation. The average response rate was about 67%.

Table 1
Monte Carlo bias, variance and mean square error(MSE) of the four point estimators and percent relative biases (R.B.) and t-statistics(t-stat) of the variance estimators based on 5,000 Monte Carlo samples

n	Method	$\hat{\theta}$			$V(\hat{\theta})$	
		Bias	Variance	MSE	R.B. (%)	t-stat
100	(PSA-MLE)	-0.01	0.0315	0.0317	-2.34	-1.12
	(PSA-CAL)	-0.01	0.0308	0.0309	-3.56	-1.70
	(AUG)	0.00	0.0252	0.0252	-0.61	-0.30
	(OPT)	0.00	0.0252	0.0252	-0.21	-0.10
400	(PSA-MLE)	-0.01	0.00737	0.00746	0.35	0.17
	(PSA-CAL)	-0.01	0.00724	0.00728	0.29	0.14
	(AUG)	0.00	0.00612	0.00612	0.07	0.03
	(OPT)	0.00	0.00612	0.00612	-0.14	-0.07

To compute the propensity score, a response model of the form

$$p(x;\phi) = \frac{\exp(\phi_0 + \phi_1 x)}{1 + \exp(\phi_0 + \phi_1 x)}$$

was postulated for parameter estimation. To obtain the augmented PSA estimator, a model for the variable of interest of the form

$$m(x;\beta) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \quad (32)$$

was postulated. Thus, model (32) is a true model under (Pop1), but it is not a true model under (Pop2).

We computed four estimators:

1. (PSA-MLE): PSA estimator in (3) using the maximum likelihood estimator of ϕ.
2. (PSA-CAL): PSA estimator in (3) with $\hat{p}_i$ satisfying the calibration constraint (15) on $(1, x)$.
3. (AUG-1): Augmented PSA estimator $\hat{\theta}^*_{\text{PSA}}$ in (19) with $\hat{\beta}$ computed by the maximum likelihood method.
4. (AUG-2): Augmented PSA estimator $\hat{\theta}^*_{\text{PSA}}$ in (19) with $\hat{\beta}$ computed by the method of Cao *et al.* (2009) discussed in Remark 1.

We considered the the augmented PSA estimator in (19) with $\hat{p}_i = p_i(\hat{\phi})$, where $\hat{\phi}$ is the maximum likelihood estimator of ϕ. The first augmented PSA estimator (AUG-1) used $\hat{m}_i = m(x_i;\hat{\beta})$ with $\hat{\beta}$ found by solving $\sum_{h=1}^{4}\sum_{i\in A_h} w_{hi}\, r_{hi}\{y_{hi} - m(x_{hi};\beta)\}(1, x_{hi}) = \mathbf{0}$, where A_h is the set of indices appearing in the sample for stratum h and w_{hi} is the sampling weight of unit i for stratum h.

Table 2 presents the simulation results for each method. In each population, the augmented PSA estimator shows some improvement comparing to the PSA estimator using the maximum likelihood estimator of ϕ or the calibration estimator of ϕ in terms of variance. Under (Pop1), since model (32) is true, there is essentially no difference between the augmented PSA estimators using different methods of estimating β. However, under (Pop2), where the assumed outcome regression model (32) is incorrect, the augmented PSA estimator with $\hat{\beta}$ computed by the method of Cao *et al.* (2009) results in slightly better efficiency, which is consistent with the theory in Remark 1. Variance estimates are approximately unbiased in all cases in the simulation study.

6. Conclusion

We have considered the problem of estimating the finite population mean of y under nonresponse using the propensity score method. The propensity score is computed from a parametric model for the response probability, and some asymptotic properties of PSA estimators are discussed. In particular, the optimal PSA estimator is derived with an additional assumption for the distribution of y. The propensity score for the optimal PSA estimator can be implemented by the augmented propensity model presented in Section 3. The resulting estimator is still consistent even when the assumed outcome regression model fails to hold.

We have restricted our attention to missing-at-random mechanisms in which the response probability depends only on the always-observed $\mathbf{x}$. If the response mechanism also depends on y, PSA estimation becomes more challenging. PSA estimation when missingness is not at random is beyond the scope of this article and will be a topic of future research.

Acknowledgements

The research was partially supported by a Cooperative Agreement between the US Department of Agriculture Natural Resources Conservation Service and Iowa State University. The authors wish to thank F. Jay Breidt, three anonymous referees, and the associate editor for their helpful comments.

Table 2
Monte Carlo bias, variance and mean square error of the four point estimators and percent relative biases (R.B.) and t-statistics of the variance estimators, based on 5,000 Monte Carlo samples

Population	Method	$\hat{\theta}_{\text{PSA}}$			$V(\hat{\theta}_{\text{PSA}})$	
		Bias	Variance	MSE	R.B. (%)	t-stat
Pop1	(PSA-MLE)	0.00	0.000750	0.000762	-1.13	-0.57
	(PSA-CAL)	0.00	0.000762	0.000769	-1.45	-0.72
	(AUG-1)	0.00	0.000745	0.000757	-1.73	-0.86
	(AUG-2)	0.00	0.000745	0.000757	-1.83	-0.91
Pop2	(PSA-MLE)	0.00	0.000824	0.000826	0.29	0.14
	(PSA-CAL)	0.00	0.000829	0.000835	-0.94	-0.46
	(AUG-1)	0.00	0.000822	0.000823	-0.71	-0.35
	(AUG-2)	0.00	0.000820	0.000821	-0.61	-0.30

References

Cao, W., Tsiatis, A.A. and Davidian, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika*, 96, 723-734.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34, 305-334.

Da Silva, D.N., and Opsomer, J.D. (2006). A kernel smoothing method of adjusting for unit non-response in sample surveys. *Canadian Journal of Statistics*, 34, 563-579.

Da Silva, D.N., and Opsomer, J.D. (2009). Nonparametric propensity weighting for survey nonresponse through local polynomial regression. *Survey Methodology*, 35, 2, 165-176.

Duncan, K.B., and Stasny, E.A. (2001). Using propensity scores to control coverage bias in telephone surveys. *Survey Methodology*, 27, 2, 121-130.

Durrant, G.B., and Skinner, C. (2006). Using missing data methods to correct for measurement error in a distribution function. *Survey Methodology*, 32, 1, 25-36.

Fay, R.E. (1992). When are inferences from multiple imputation valid? *Proceedings of the Survey Research Methods Section*, American Statistical Association, 227-232.

Folsom, R.E. (1991). Exponential and logistic weight adjustments for sampling and nonresponse error reduction. *Proceedings of the Social Statistics Section*, American Statistical Association, 197-202.

Fuller, W.A., Loughin, M.M. and Baker, H.D. (1994). Regression weighting in the presence of nonresponse with application to the 1987-1988 Nationwide Food Consumption Survey. *Survey Methodology*, 20, 1, 75-85.

Iannacchione, V.G., Milne, J.G. and Folsom, R.E. (1991). Response probability weight adjustments using logistic regression. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 637-642.

Isaki, C.T., and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77, 89-96.

Kim, J.K. (2004). Finite sample properties of multiple imputation estimators. *The Annals of Statistics*, 32, 766-783.

Kim, J.K., and Kim, J.J. (2007). Nonresponse weighting adjustment using estimated response probability. *Canadian Journal of Statistics*, 35, 501-514.

Kim, J.K., Navarro, A. and Fuller, W.A. (2006). Replication variance estimation for two-phase stratified sampling. *Journal of the American Statistical Association*, 101, 312-320.

Kim, J.K., and Rao, J.N.K. (2009). A unified approach to linearization variance estimation from survey data after imputation for item nonresponse. *Biometrika*, 96, 917-932.

Kott, P.S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, 32, 2, 133-142.

Lee, S. (2006). Propensity score adjustment as a weighting scheme for volunteer panel web surveys. *Journal of Official Statistics*, 22, 329-349.

Little, R.J.A. (1988). Missing-data adjustments in large surveys. *Journal of Business and Economic Statistics*, 6, 287-296.

Nevo, A. (2003). Using weights to adjust for sample selection when auxiliary information is available. *Journal of Business and Economic Statistics*, 21, 43-52.

Pfeffermann, D., Krieger, A.M. and Rinott, Y. (1998). Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica*, 8, 1087-1114.

Randles, R.H. (1982). On the asymptotic normality of statistics with estimated parameters. *The Annals of Statistics*, 10, 462-474.

Rizzo, L., Kalton, G. and Brick, J.M. (1996). A comparison of some weighting adjustment methods for panel nonresponse. *Survey Methodology*, 22, 1, 43-53.

Robins, J.M., Rotnitzky, A. and Zhao, L.P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89, 846-866.

Rosenbaum, P.R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association*, 82, 387-394.

Rosenbaum, P.R., and Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.

Shao, J., and Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association*, 94, 254-265.

Singh, A.C., and Folsom, R.E. (2000). Bias corrected estimating function approach for variance estimation adjusted for poststratification. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 610-615.

Wu, C., and Sitter, R.R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, 96, 185-193.